Стаття присвячена огляду підходів до аналітичного моделювання процесів, що мають місце у чергах телекомунікаційного вузла у випадках, коли тривалість обслуговування має розподіл з «важким хвостом»

Ключові слова: СМО, СКВ, функція розподілу

Статья посвящена рассмотрению подходов к аналитическому моделированию процессов, происходящих в очередях телекоммуникационного узла в случаях, когда длительность обслуживания имеет распределение с «тяжелым хвостом»

Ключевые слова: СМО, СКО, функция распределения

Article is dedicated to overview approaches of queueing processes analytical modeling in cases when service duration has «heavy tail» distribution

Key words: QS, MSD, distribution function

УДК 261.391

АНАЛИЗ ПРОИЗВОДИТЕЛЬНОСТИ УЗЛА СЕТИ В УСЛОВИЯХ НЕСТАЦИОНАРНОГО ТРАФИКА

Ю.В. Андрушко
Аспирант
Кафедра телекоммуникационных систем
Харьковский национальный университет радиоэлектроники
пр. Ленина 14, г. Харьков, Украина, 61000
Контактный тел.: (057) 702-13-20
E-mail: ya@kture.kharkov.ua

1. Введение

Неотъемлемой составляющей проектирования, эксплуатации и модернизации телекоммуникационных систем является аналитическое моделирование процесса их функционирования. Для этого активно используются модели и методы теории массового обслуживания [1 – 2].

На вход системы массового обслуживания (СМО) с накопителем конечной емкости m, $1 \le m < \infty$. поступает пуассоновский поток заявок интенсивности λ. Длительности обслуживания заявок имеют гамма-распределение с плотностью распределения вида [3]:

$$B(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}},$$

или эквивалентно:

$$F(x) = \frac{\upsilon^{\alpha} x^{\alpha-1} e^{\upsilon x}}{\Gamma(\alpha)}, x \ge 0, \tag{1}$$

где $\upsilon = \alpha\mu$;

$\mu^{-1} = t_{cp}$ - среднее время обслуживания;

$\alpha = 1/\sigma^2$, σ - СКО времени обслуживания.

Если при переполнении накопителя поступающая на систему заявка теряется и вновь не возобновляется, а так же $t_{cp} < \infty$ и $\sigma < \infty$, такая СМО в классификации Кендалла может быть описана как $M/\Gamma/1/m$.

В рамках задачи аналитического моделирования процессов в телекоммуникационных системах особого внимания заслуживает моделирование процессов, происходящих в очередях телекоммуникационного узла. Большое внимание при моделировании этих процессов уделяется ситуациям, когда длительности обслуживания имеют распределение с так называемыми «тяжелыми хвостами» [2]. Основное свойство распределения такого типа заключается в том, что его дополнительная функция распределения (ФР) медленно убывает при $t \to \infty$. Распределения с «тяжелыми хвостами» также характеризуются тем, что имеют большие или неограниченные дисперсии. В рамках данного раздела будут проанализированы свойства процесса ограниченной очереди при пуассоновском или марковском входящем потоке заявок и гамма-распределении времени их обслуживания в условиях, когда среднеквадратическое отклонение (СКО) времени обслуживания неограниченно растет.

Несмотря на то, что гамма-распределение не является распределением с «тяжелым хвостом», однако при стремлении СКО (при фиксированном среднем значении) к бесконечности также неограниченно растет и дисперсия. Это позволяет предположить что результаты для процесса очереди, полученные в условиях неограниченного роста СКО, могут в той или иной мере отражать поведение характеристик очереди для случая ФР с «тяжелыми хвостами».

2. Аналитические выражения для сетевых характеристик

Если обозначить через p_k стационарную вероятность того, что в системе $M/\Gamma/1/m$ в произвольный момент времени находится k заявок, для вычисления стационарного распределения вероятностей $\{p_k, k = \overline{1,M}\}$ ($M = m + 1$), можно использовать следующие формулы:

$$p_1 = \frac{q_1}{\lambda} - p_0, p_k = \frac{q_k}{\lambda}, k = \overline{2,m}, \tag{2}$$

$$p_M = 1 - \frac{q_{\cdot}}{\lambda}. \tag{3}$$

Интенсивность выхода обслуженных заявок λ_D:

$$\lambda_D = \mu(1-p_0). \tag{4}$$

Рассмотрим моменты распределения. Если плотность распределения времени обслуживания определяется формулой (1), то величины $\beta_k \equiv \beta_k(\lambda)$ вычисляются по следующим рекуррентным формулам:

$$\beta_0 = \left(\frac{\nu}{\lambda+\nu}\right)^{\alpha}, \beta_k = \frac{k+\alpha-1}{k}\frac{\lambda}{\lambda+\nu}\beta_{k-1}, k \ge 1. \tag{5}$$

Рассмотрим ситуацию, когда $\sigma \to \infty$. В этом случае, нельзя утверждать, что вероятности $\{p_k, k=\overline{0,M}\}$ задают стационарное распределение вероятностей, так как формулы получены в предположении, что $t_{ср}$ и σ конечны. Однако, исходя из них, можно получить выражения для некоторых показателей производительности СМО в условиях, когда $\sigma \to \infty$. Далее, предельные значения (при $\sigma \to \infty$) исследуемых величин обозначены «*».

Выражение для β_0 в (5) приводится к виду:

$$\beta_0 = \left(\frac{1}{1+\rho\sigma^2}\right)^{\frac{1}{\sigma^2}}, \tag{6}$$

где $\rho = \lambda/\mu = \lambda t_{ср}$.

Переходя в (6) к пределу при $\sigma \to \infty$, с учетом (5) получим:

$$\beta_k^* = \lim_{\sigma\to\infty} \beta_k = 0, k \ge 1. \tag{7}$$

Исходя из (3) при $\sigma \to \infty$:

$$\tilde{q}_1^* = \lambda, \tilde{q}_i^* = 0, i = \overline{2,m}. \tag{8}$$

Таким образом предельные значение для вероятностей:

$$p_0^* = \frac{1}{1+\rho}, \tag{9}$$

$$p_k^0 = 0, k = \overline{1,m}, \tag{10}$$

$$p_M^* = \frac{\rho}{1+\rho}. \tag{11}$$

Проанализировав (9) – (11) можно сделать вывод, что при $\sigma \to \infty$ СМО $M/\Gamma/1/m$ функционирует как система с двумя состояниями: состоянием (0), отвечающим ситуации, когда система $M/M/1/0$ полностью свободна, и состоянием (M), соответствующим случаю, когда система занята обслуживанием и в ней находится M заявок.

Используя полученные формулы, можно получить предельные выражения для ряда других показателей производительности СМО.

Из (9) – (11) для среднего числа заявок в системе N^*:

$$N^* = Mp_M^* = \frac{M\rho}{1+\rho}. \tag{12}$$

Для средней длины очереди Q^*:

$$Q^* = mrp_M^* = \frac{mp}{1+\rho}. \tag{13}$$

Для интенсивности обслуженного потока – интенсивности выхода λ_D^* из (4) и (9):

$$\lambda_D^* = \frac{\lambda}{1+\rho}. \tag{14}$$

Вероятности потерь π^* определяется как:

$$\pi^* = p_M^* = \frac{\rho}{1+\rho}. \tag{15}$$

Приняв во внимание формулы Литтла ($\lambda(1-\pi)w = Q$, $\lambda(1-\pi)v = N$), которые описывают связь средней длины очереди Q со средним временем ожидания w и среднее число заявок в системе N со средним временем пребывания заявки в системе v, для предельных значений w^* и v^* можно получить:

$$w^* = \frac{m}{\mu}, \tag{16}$$

$$v^* = \frac{M}{\mu}. \tag{17}$$

Основываясь на формулах (9) – (13), получим выражения для значений среднеквадратичных отклонений длины очереди σ_Q^* и числа заявок в системе σ_N^*:

$$\sigma_Q^* = \frac{m\sqrt{\rho}}{1+\rho}, \tag{18}$$

$$\sigma_N^* = \frac{M\sqrt{\rho}}{1+\rho}. \tag{19}$$

Для получения выражения предельных значений показателей производительности, примем $\rho = 1$:

$$p_0^* = p_M^* = \pi^* = \frac{1}{2},$$
$$Q^* = \frac{m}{2},\ N^* = \frac{M}{2},\ \lambda_D^* = \frac{\lambda}{2}, \tag{20}$$
$$\sigma_Q^* = \frac{m}{2},\ \sigma_N^* = \frac{M}{2}.$$

Если на вход рассмотренной системы поступает Марковский поток, который характеризуется матрицами Λ и N порядка l. Элемент Λ_{ij} матрицы Λ есть интенсивность перехода процесса генерации заявки с фазы i на фазу j без поступления новой заявки, а элемент N_{ij} матрицы N есть также интенсивность перехода процесса генерации заявки с фазы i на фазу j, но с поступлением новой заявки. Предполагается, что матрица N отлична от нулевой матрицы, а матрица $\tilde{\Lambda} = \Lambda + N$ неразложима.

Можно обозначить стационарную вероятность того, что в произвольный момент времени в системе нет заявок, а процесс генерации заявок находится на фазе i как p_{i0}, в таком случае p_{ik} будет представлять стационарную вероятность того, что в произвольный момент времени в системе находится k заявок, а генерируемая заявка проходит фазу i. Введем векторы $P_л = (p_{1k}, p_{2k}, ..., p_{lk})^T, k = \overline{0,M}$. Для вычисления стационарного распределения вероятностей $\{P_k, k=\overline{0,M}\}$ состояний системы $MAP/G/1/m$ можно использовать алгоритм, изложенный [4].

Для СМО $MAP/\Gamma/1/m$ плотность распределения времени обслуживания определяется формулой (1). Прежде всего можно отметить, что выражения для предельных значений экспоненциальных моментов $\beta_л^*$

не зависят от параметра λ. Поэтому для вычисления предельных (при $\sigma \to \infty$) значений вероятностей состояний системы $MAP/\Gamma/1/m$ можно использовать (6) и (7) для вычисления предельных значений $\beta_0^*(a)$ и $\beta_0^*(\tilde{a})$:

$$\beta_0^*(a) = \beta_0^* = 1, \;\; \beta_0^*(a) = \beta_k^* = 0, \;\; k \geq 1;$$

$$\beta_0^*(\tilde{a}) = \beta_0^* = 1, \;\; \beta_0^*(\tilde{a}) = \beta_k^* = 0, \;\; k \geq 1. \qquad (21)$$

Используя полученные ранее результаты, можно показать, что:

$$C = \frac{1}{1+\rho},$$

и, следовательно,

$$P_0^* = \pi \frac{1}{1+\rho}, \qquad (22)$$

$$P_K^* \pi \frac{\rho}{1+\rho}. \qquad (23)$$

Вероятности $p_0^* = P_0^{*T} 1$ и $p_M^* = P_M^{*T} 1$ определяются формулами:

$$p_0^* = \frac{1}{1+\rho}, \qquad (24)$$

$$p_M^* \frac{\rho}{1+\rho}, \qquad (25)$$

т.е. p_0^* и p_M^* совпадают с аналогичными вероятностями для СМО $MAP/\Gamma/1/m$.

Вероятность потери π для СМО $MAP/G/1/m$ определяется формулой [4]:

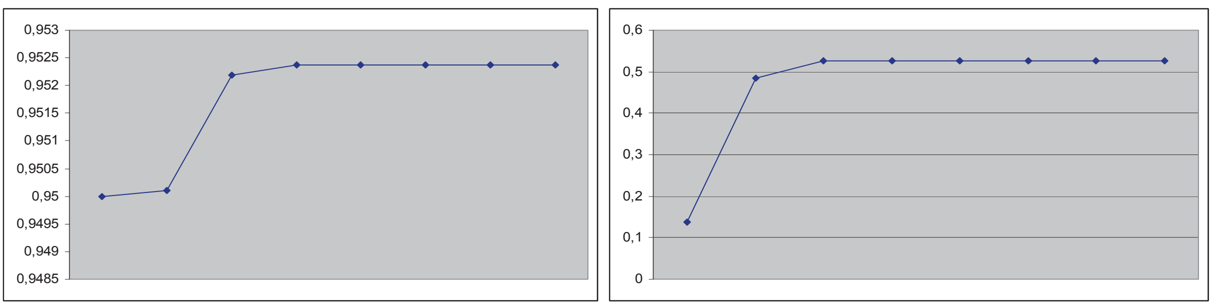
$$\pi = \frac{1}{\lambda} P_K^T \vec{\lambda}. \qquad (26)$$

Предельное значение вероятности потерь π^* для СМО $MAP/\Gamma/1/r$ вычисляется по той же формуле, что и для $M/\Gamma/1/m$:

$$\pi^* = p_M^* = \frac{\rho}{1+\rho}. \qquad (27)$$

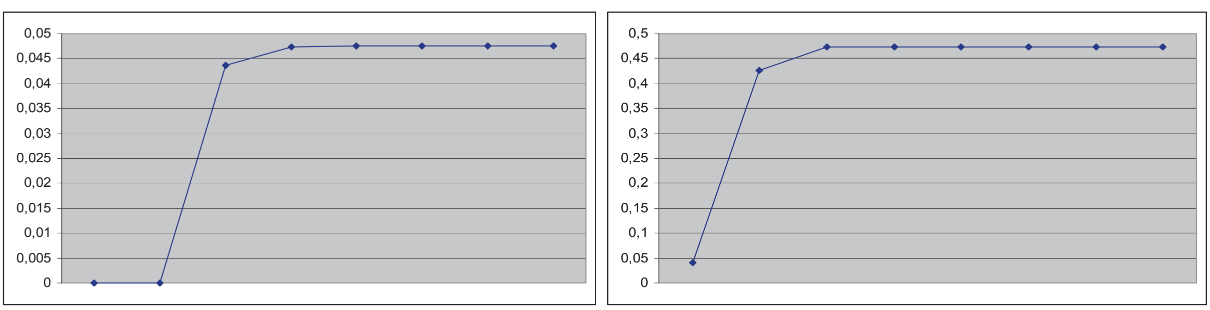
3. Численный анализ результатов

Далее приводятся результаты, полученные для СМО $M/\Gamma/1/m$, иллюстрирующие сходимость по-



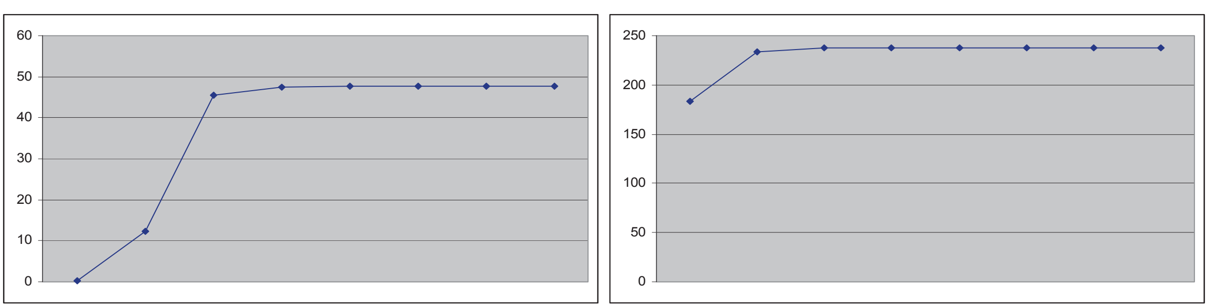
а) б)

Рис. 1. Вероятность простоя системы p_0



а) б)

Рис. 2. Вероятность потери заявки π

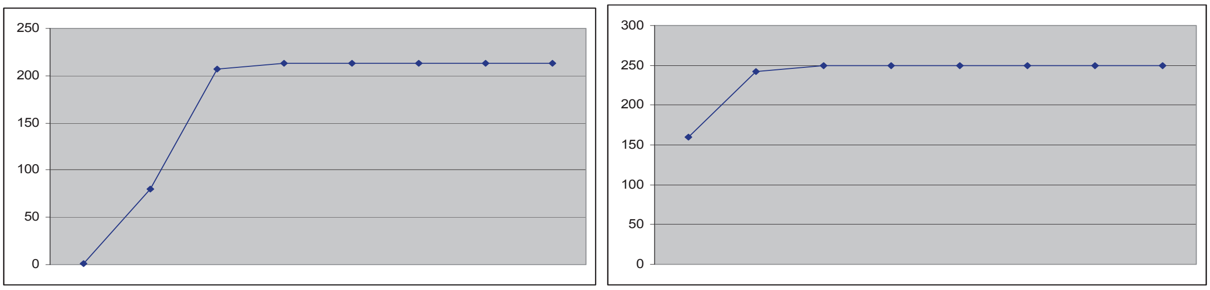


а) б)

Рис. 3. Среднее число заявок в системе N

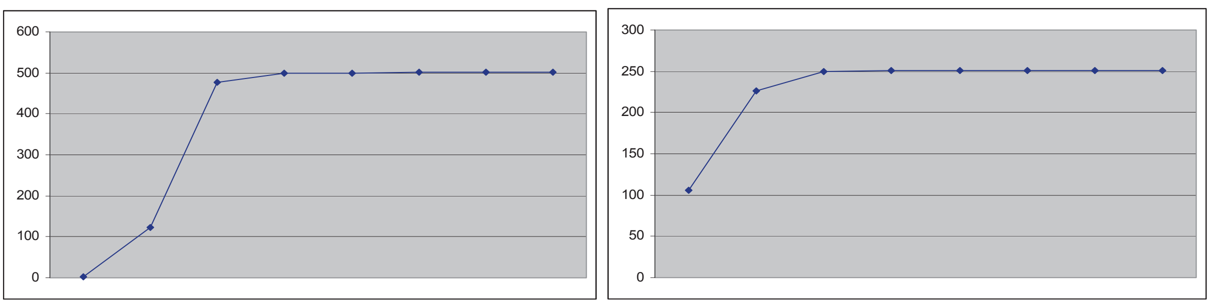
казателей производительности системы при $\sigma \to \infty$ к их предельным значениям. На рис. 1 – 5 (в логарифмическом масштабе) приведены зависимости различных сетевых параметров от среднеквадратического отклонения времени обслуживания. Рис. 1 – 5 а) –иллюстрируют значения сетевых параметров при $\lambda = 0{,}1$, $\mu = 2$ и $m = 1000$ для малых $(\rho = 0{,}05)$, а рис. 1 – 5 б) – при $\lambda = 1{,}8$, $\mu = 2$ и $m = 500$ для больших $(\rho = 0{,}9)$ нагрузок.

ными ($\sigma \to \infty$) значениями. Это также подтверждается расчетами, выполненными по формулам для (теоретических) предельных значений данных показателей. Следует заметить, что приведенные показатели производительности отражают поведение СМО в условиях малой загрузки $(\rho = 0{,}05)$. Однако, как следует из графиков, при больших значениях СКО $(\sigma \geq 1000)$ вероятность потери заявки достаточно велика $(\pi > 0{,}04)$, что при значении емкости



а) б)

Рис. 4. Среднеквадратическое отклонение числа заявок в системе σ_n



а) б)

Рис. 5. Среднее время пребывания заявки в системе ν

4. Выводы

В статье проведен анализ подходов к аналитическому описанию процессов очередей телекоммуникационного узла. Основное внимание при этом уделялось случаю, когда длительности обслуживания имеют распределение с так называемыми «тяжелыми хвостами». Был затронут ряд фундаментальных свойств и характеристик таких распределений, а так же показано, что, несмотря на то, что гамма-распределение не является классическим примером распределения с «тяжелым хвостом», в случаях когда СКО времени обслуживания заявки в узле $\sigma \to \infty$, гамма-распределение приобретает свойства таких распределений. Используя известный математический аппарат для гамма-распределения, были получены формулы для нахождения значений важнейших показателей производительности сети для различных СМО. Кроме прочего, отдельно были затронуты ситуации с варьированием параметров загрузки и времени обслуживания заявки. В заключительном параграфе статьи был проведен численный анализ полученных результатов. На рис. 1 – 5 а) показано, что значения показателей производительности при $\sigma = 10^5$ и $\sigma = 10^6$ (два последних отсчета) для заданной точности вычислений совпадают, т.е. эти значения фактически можно считать предельными ($\sigma \to \infty$) значениями.

накопителя m = 1000 порождает вопросы к достоверности. Также значительны среднее число заявок в системе, равное примерно 47 (с учетом среднеквадратичного отклонения число заявок в системе может быть более 260), и среднее время пребывания заявки в системе, которое при столь малой загрузке системы превышает 500 единиц времени.

Тем не менее, эти факты не кажутся неправдоподобными, если принять во внимание, что СМО рассматривается в условиях большого СКО времени обслуживания, что при фиксированном среднем времени обслуживания эквивалентно большой его дисперсии.

По результатам второго моделирования, СМО исследуется в условиях большой загрузки $(\rho = 0{,}9)$. Так же как и для малой загрузки, при $\sigma \geq 10^5$ значения показателей производительности, приведенные в двух предпоследних отсчетах, можно считать (при заданной точности вычислений) совпадающими с их предельными значениями, что также подтверждается расчетами для (теоретических) предельных значений этих показателей, которые представлены последним отсчетом на графиках. Для большой загрузки также наблюдается эффект значительного увеличения вероятности потерь с ростом СКО σ, что опять же объясняется большой дисперсией времени обслуживания.

Литература

1. Башарин Г.П. Анализ очередей в вычислительных сетях. Теория и методы расчета / Г.П. Башарин, П.П. Бочаров, Я.А. Коган – М.: Наука. Гл. ред. физ.-мат. лит., 1989. - 336 с.
2. Шелухин О.И. Фрактальные процессы в телекоммуникациях : учеб, пособие / О.И. Шелухин, А.М. Тенякшев, А.В. Осин ; Под ред. Шелухина О.И. М.: Радиотехника, 2003. – 479с.
3. Абрамовиц, М. Справочник по специальным функциям с формулами, графиками и математическими таблицами : пер. с англ. - М.: Наука, 1979. – 835с.
4. Бочаров П.П Теория массового обслуживания : учеб, пособие / П.П Бочаров, А.В. Печинкин М.: Изд-во РУДН, 1995. – 529с.

УДК 001.891:65.011.56

ПОДХОД К КЛАССИФИКАЦИИ ПОЛЬЗОВАТЕЛЕЙ В СОЦИАЛЬНЫХ СЕТЯХ

А.А. Куликова
Кафедра искусственного интеллекта
Харьковский национальный университет радиоэлектроники
пр. Ленина,14, г. Харьков, Украина, 61166
Контактный тел.: (057) 337-27-53, 093-776-41-20
E-mail: ganna.kulikova@gmail.com

У статті розглянуто проблему спаму в соціальних мережах, класифікація легітимних і нелегітимних користувачів

Ключові слова: соціальні мережі, класифікація, спам, алгоритм

В статье рассмотрена проблема спама в социальных сетях, классификация легитимных и нелегитимных пользователей

Ключевые слова: социальные сети, классификация, спам, алгоритм

The problem of spam in social networks, legitimate and non-legitimate users' classification is considered in this article

Key words: social networks, classification, spam, algorithm

1. Введение

Социальные сети, привлекающие сегодня к себе всеобщее внимание пользователей Интернета, сформировались за очень короткий промежуток времени. Они объединяют в себе блоги (сетевые дневники), сети медиа-ресурсов, сети персональной информации (MySpace, LinkedIn, Facebook, Вконтакте), системы закладок (del.icio.us), wiki-энциклопедии и другие. Данные Web-сайты представляют собой автоматизированную социальную среду для обеспечения коммуникации как отдельных, так и групп пользователей, объединенных общими интересами. Количество пользователей в этих сетях увеличивается с беспрецедентной скоростью, вызывая интерес у представителей науки, бизнеса и IT-индустрии [1]. Такие Web-сайты фактически представляют собой большое хранилище общедоступной информации, в первую очередь, персонального характера.

Однако в то же время развитие Internet, технологий проектирования социальных сетей привело к тому, что одной из основных проблем пользователей стал избыток информации, в том числе и незапрошенной, – спама.

Спам представляет собой масштабную рассылку коммерческой, политической и иной рекламы (информации) или иного вида сообщений лицам, не выражавшим желания их получать. Значительная часть атак основана на методах социальной инженерии (привлечение пользователей недобросовестной рекламой и т.д.), другая – на использовании уязвимостей в механизмах работы социальных сетей. Существует достаточно много видов спама, распространяемого в социальных сетях, но, прежде всего, стоит отметить рекламу, некоторые виды мошенничества, фишинг, распространение вредоносного программного обеспечения.

Технологии рассылки спама в социальных сетях совершенствуются: спаммеры отмечают пользователей социальной сети на фотографиях, видеозаписях, добавляются в друзья, приглашают в группы и так далее, в целом используют все возможности социальной сети в корыстных целях.

Борьба со спаммерами в социальных сетях важна для улучшения сервисов, предоставляемых социальной сетью для участников, уменьшения количества нежелательного и опасного контента, а так же для развития самих социальных сетей.